Songklanakarin J. Sci. Technol.
38 (6), 645-656, Nov. - Dec. 2016

http://www.sjst.psu.ac.th

Original Article

The Poisson-generalised Lindley distribution and its applications

Weerinrada Wongrin and Winai Bodhisuwan*

Department of Statistics, Faculty of Science, Kasetsart University, Chatuchak, Bangkok, 10900 Thailand.

Received: 30 October 2015; Accepted: 25 February 2016

Abstract

The Poisson distribution plays important role in count data analysis. However, the Poisson distribution cannot model some data with over-dispersion because of its property, equi-dispersion. Here we propose a new distribution for over-dispersed count data, namely the Poisson-generalised Lindley distribution. Basic properties of the distribution and special cases are also derived. In addition, the new distribution is applied to some real data sets using the method of maximum likelihood for parameter estimation. The results based on *p*-value of the discrete Anderson-Daring test show that the new distribution can be used as an alternative model for count data analysis.

Keywords: count data, mixed Poisson distribution, generalised Lindley distribution, over-dispersion

1. Introduction

Count data are used to describe many phenomena such as the insurance claim numbers, number of yeast cells, number of chromosomes, etc. (Panjer, 2006). Count data analysis can use a Poisson distribution to describe the data if its variance to mean ratio, called the dispersion index, is unity (equi-dispersion) (Johnson *et al.*, 2005). However, many practical count data sets do not satify the equi-dispersion assumption. Therefore, the Poisson distribution is inflexible to model many count data sets (Raghavachari *et al.*, 1997; Karlis and Xekalaki, 2005). An inequality of variance and mean is called over-dispersion if the variance exceeds the mean, and under-dispersion if the variance is less than the mean.

Many researchers have looked at the over-dispersion issue which can be addressed by the use of mixed Poisson distributions (Raghavachari *et al.*, 1997; Karlis and Xekalaki, 2005; Panjer, 2006). Mixed Poisson distributions arise when the mean of the Poisson is a random variable with some specified distribution. The distribution of the Poisson rate is the so-called mixing distribution (Everitt and Hand, 1981; Raghavachari *et al.*, 1997).

The negative binomial (NB) distribution, which is a traditional mixed Poisson distribution where the mean of the Poisson variable is distributed as a gamma random variable, was derived by Greenwood and Yule (1920). It has increasingly become a popular alternative distribution to the Poisson distribution. However, the NB distribution may not be appropriate for some over-dispersed count data.

Other mixed Poisson distributions arise from alternative mixing distributions. If the mean of the Poisson follows an inverse Gaussian, resulting in a Poisson-inverse Gaussian (Holla, 1967). The Poisson-Lindley (PL) (Sankaran, 1970) and generalised Poisson - Lindley (Mahmoudi and Zakerzadeh, 2010) distributions were obtained where the mixing distributions are the Lindley and the generalised Lindley distributions, respectively. Recently, a Poisson-weighted exponential distribution was developed by Zamani *et al.* (2014), where a weighted exponential is the mixing distribution.

It has been found that the general characteristics of the mixed Poisson distribution follow some characteristics of its mixing distribution. Depending on the choice of the mixing distribution, various mixed Poisson distributions have been

* Corresponding author.
Email address: fsciwnb@ku.ac.th

constructed. However, since their mathematical forms are often complicated, only a few of them have been applied in practice. Furthermore, in any case, there are naturally situations where a good fit is not obtainable with existing developed distributions (Karlis and Xekalaki, 2005).

The purpose of this paper is to present an alternative distribution for over-dispersed count data, namely the Poisson-generalised Lindley (PGL) distribution. It is obtained by mixing the Poisson distribution with a new generalised Lindley (NGL) distribution (Elbatal *et al.*, 2013).

The probability density function of the three-parameter NGL distribution, which generalised the Lindley distribution, has many shapes. Due to its flexible shape, it can be used as an alternative model for fitting positive real-valued data in many areas. For this paper, we show that the proposed mixed Poisson distribution is suitable for modelling real count data in some situations.

In Section 2, the new distribution, called the PGL distribution, is introduced. Some special cases of the distribution are also considered in this section. Its basic mathematical properties including the moment generating function, probability generating function and moments are derived in Section 3. We also discuss the method of parameter estimation in Section 4. Finally, applications of the PGL to real data are given in Section 5.

2. The Poisson-Generalised Lindley Distribution

Let Y be the random variable that represents the total number of outcomes of a particular experiment. The simple model for the probabilities of the possible outcomes of this experiment is the Poisson distribution, with probability mass function (pmf)

$$p(y)=\frac{\exp(-\lambda)\lambda^{y}}{y!};\quad y=0,1,2,\ldots,\text{ and }\lambda>0. \tag{1}$$

An important property of the Poisson distribution is that the positive real number λ equals both the expected value of Y and its variance, i.e. $\mathrm{E}(Y)=\mathrm{Var}(Y)=\lambda$.

In 2013, the NGL distribution was introduced (Elbatal *et al.*, 2013). It is a three-parameter continuous distribution used to analyse lifetime data. It can model many shapes of hazard rate function. The probability density function (pdf) can be obtained by concept of the finite mixture distribution

$$\begin{aligned}g(\lambda)&=pg_1(\lambda)+(1-p)g_2(\lambda)\\&=\frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}\lambda^{\alpha-1}}{\Gamma(\alpha)}+\frac{\theta^{\beta}\lambda^{\beta-1}}{\Gamma(\beta)}\right)\exp(-\theta\lambda);\\&\lambda>0,\text{ and }\alpha,\beta,\theta>0,\end{aligned} \tag{2}$$

where $\lambda_1\sim\mathrm{Gamma}(\alpha,\theta)$, $\lambda_2\sim\mathrm{Gamma}(\beta,\theta)$, $p=\theta/(\theta+1)$, and the gamma function is defined as $\Gamma(t)=\int_0^\infty x^{t-1}\exp(-x)\mathrm{d}x$.

The PGL distribution is a new mixed Poisson distribution. It is obtained by mixing the Poisson distribution with the NGL distribution. We provide a general definition of this distribution which will subsequently expose its pmf.

Definition 1:

Let $Y\mid\lambda$ be a random variable following a Poisson distribution with parameter λ, $Y\mid\lambda\sim\mathrm{Pois}(\lambda)$. If λ is distributed as a new generalised Lindley with parameters α, β and θ, denoted by $\lambda\sim\mathrm{NGL}(\alpha,\beta,\theta)$, then Y is called a Poisson-generalised Lindley variable.

Proposition 1:

Let Y be a random variable according to the PGL probability function, denoted by $Y\sim\mathrm{PGL}(\alpha,\beta,\theta)$, the pmf of Y is

$$p(y;\alpha,\beta,\theta)=\frac{1}{y!(\theta+1)^{y+1}}\left(\left(\frac{\theta}{\theta+1}\right)^{\alpha}\frac{\theta\,\Gamma(y+\alpha)}{\Gamma(\alpha)}+\left(\frac{\theta}{\theta+1}\right)^{\beta}\frac{\Gamma(y+\beta)}{\Gamma(\beta)}\right), \tag{3}$$

for $y=0,1,2,\ldots,$ and $\alpha,\beta,\theta>0$.

Proof:

Since $Y\mid\lambda\sim\mathrm{Pois}(\lambda)$ and $\lambda\sim\mathrm{NGL}(\alpha,\beta,\theta)$, the marginal pmf of $Y\sim\mathrm{PGL}(\alpha,\beta,\theta)$ can be obtained by

$$p(y;\alpha,\beta,\theta)=\int_0^\infty p(y)g(\lambda)\mathrm{d}\lambda. \tag{4}$$

By substituting Eq. (1) and Eq. (2) into Eq. (4), we derive the marginal pmf of the PGL distribution:

$$\begin{aligned}p(y;\alpha,\beta,\theta)&=\int_0^\infty\frac{\exp(-\lambda)\lambda^{y}}{y!}\frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}\lambda^{\alpha-1}}{\Gamma(\alpha)}+\frac{\theta^{\beta}\lambda^{\beta-1}}{\Gamma(\beta)}\right)\exp(-\theta\lambda)\mathrm{d}\lambda\\&=\frac{1}{y!(\theta+1)}\left(\frac{\theta^{\alpha+1}}{\Gamma(\alpha)}\int_0^\infty\lambda^{y+\alpha-1}\exp(-(\theta+1)\lambda)\mathrm{d}\lambda\right.\\&\qquad\left.+\frac{\theta^{\beta}}{\Gamma(\beta)}\int_0^\infty\lambda^{y+\beta-1}\exp(-(\theta+1)\lambda)\mathrm{d}\lambda\right)\\&=\frac{1}{y!(\theta+1)^{y+1}}\left(\frac{\theta^{\alpha+1}\Gamma(y+\alpha)}{(\theta+1)^{\alpha}\Gamma(\alpha)}+\frac{\Gamma(y+\beta)}{\Gamma(\beta)}\left(\frac{\theta}{\theta+1}\right)^{\beta}\right)\\&=\frac{1}{y!(\theta+1)^{y+1}}\left(\left(\frac{\theta}{\theta+1}\right)^{\alpha}\frac{\theta\Gamma(y+\alpha)}{\Gamma(\alpha)}+\left(\frac{\theta}{\theta+1}\right)^{\beta}\frac{\Gamma(y+\beta)}{\Gamma(\beta)}\right).\end{aligned}$$

Moreover, if a random variable corresponding to $y\in\mathbb{Z}^*$, $\mathbb{Z}^*=\{0,1,2,\ldots\}$ is a sample space, the pmf of Y is the probability function with the following properties:

I. If a random variable Y is distributed as the PGL with the pmf in Eq. (3), when $y = 0$, we obtain

$$P(Y=0)=\frac{1}{(\theta+1)}\left(\left(\frac{\theta}{\theta+1}\right)^{\alpha}\frac{\theta\Gamma(\alpha)}{\Gamma(\alpha)}+\left(\frac{\theta}{\theta+1}\right)^{\beta}\frac{\Gamma(\beta)}{\Gamma(\beta)}\right)$$

$$=\left(\frac{\theta}{\theta+1}\right)^{\alpha+1}+\frac{\theta^{\beta}}{\left(\theta+1\right)^{\beta+1}},$$

for $\alpha,\beta,\theta>0$, $P(Y=0)\geq 0$.

If $y=1$, we have

$$P(Y=1)$$

$$=\frac{1}{(\theta+1)^{2}}\left(\left(\frac{\theta}{\theta+1}\right)^{\alpha}\frac{\theta\Gamma(\alpha+1)}{\Gamma(\alpha)}+\left(\frac{\theta}{\theta+1}\right)^{\beta}\frac{\Gamma(\beta+1)}{\Gamma(\beta)}\right)$$

$$=\frac{1}{(\theta+1)^{2}}\left(\left(\frac{\theta}{\theta+1}\right)^{\alpha}\frac{\theta\Gamma(\alpha+1)}{\Gamma(\alpha)}+\left(\frac{\theta}{\theta+1}\right)^{\beta}\frac{\Gamma(\beta+1)}{\Gamma(\beta)}\right)$$

$$=\frac{\theta^{\alpha+1}}{\left(\theta+1\right)^{\alpha+2}}\alpha+\frac{\theta^{\beta}}{\left(\theta+1\right)^{\beta+2}}\beta,$$

where $\alpha,\beta,\theta>0$, then $P(Y=1)\geq 0$.

In the same manner, it is obviously, $y=0,1,2,3,\ldots$, the probability of Y is greater than or equal to zero. Therefore, $P(Y=y)$ sastifies $P(Y=y)\geq 0$, for all $Y\in\mathbb{Z}^{*}$.

II. If a random variable y is distributed as the PGL with the pmf in Eq. (3), then

$$\sum_{\forall y}p(y;\alpha,\beta,\theta)$$

$$=\sum_{\forall y}\frac{1}{y!(\theta+1)^{y+1}}\left(\left(\frac{\theta}{\theta+1}\right)^{\alpha}\frac{\theta\Gamma(y+\alpha)}{\Gamma(\alpha)}+\left(\frac{\theta}{\theta+1}\right)^{\beta}\frac{\Gamma(y+\beta)}{\Gamma(\beta)}\right)$$

$$=\sum_{y=0}^{\infty}\frac{\theta^{\alpha+1}\Gamma(y+\alpha)}{y!\Gamma(\alpha)(\theta+1)^{\alpha+y+1}}+\sum_{y=0}^{\infty}\frac{\theta^{\beta}\Gamma(y+\beta)}{y!\Gamma(\beta)(\theta+1)^{\beta+y+1}}$$

$$=\frac{\theta^{\alpha+1}\Gamma(\alpha)}{\Gamma(\alpha)(\theta+1)^{\alpha+1}}+\frac{\theta^{\alpha+1}\Gamma(1+\alpha)}{\Gamma(\alpha)(\theta+1)^{\alpha+2}}+\frac{\theta^{\alpha+1}\Gamma(2+\alpha)}{2\Gamma(\alpha)(\theta+1)^{\alpha+3}}$$

$$+\cdots+\frac{\theta^{\beta}\Gamma(\beta)}{\Gamma(\beta)(\theta+1)^{\beta+1}}+\frac{\theta^{\beta}\Gamma(1+\beta)}{\Gamma(\beta)(\theta+1)^{\beta+2}}$$

$$+\frac{\theta^{\beta}\Gamma(2+\beta)}{2\Gamma(\beta)(\theta+1)^{\beta+3}}+\cdots$$

$$=\theta^{\alpha+1}\left(\frac{\theta+1}{\theta}\right)^{\alpha}\left(\frac{1}{\theta+1}\right)^{\alpha+1}+\theta^{\beta}\left(\frac{\theta+1}{\theta}\right)^{\beta}\left(\frac{1}{\theta+1}\right)^{\beta+1}$$

$$=\left(\frac{\theta}{\theta+1}\right)+\left(\frac{1}{\theta+1}\right)=1.$$

Hence, $P(\mathbb{Z}^{*})=1$.

From I and II, it can verify that the pmf of $Y\mid\alpha,\beta,\theta$ is a probability function.

Figure 1 illustrates pmf plots of the PGL distribution for some selected parameter values. It was found that the shape of the PGL distribution is characterised by long-tailed behaviour and also that the distribution has the same shape as the NGL distribution with appropriate parameter values. The parameters α and β are the shape parameters and θ is the rate parameter of the PGL distribution. Furthermore, the PGL is a bimodal distribution when parameters α and β are very different for appropriate values of the parameter θ as shown in Figure 2.

2.1 Special cases

This section presents some special cases of the PGL distribution.

Corollary 1:

For $\alpha=1,\beta=2$, we obtain the PL distribution with the pmf

$$f(y;\theta)=\frac{\theta^{2}(y+\theta+2)}{(\theta+1)^{y+3}}. \tag{5}$$

The PL distribution is a mixed Poisson distribution, which is a well-known discrete distribution. It has been used previously to model count data (Sankaran, 1970, Shanker and Fesshaye, 2015).

Corollary 2:

For $\alpha=\beta=r$, we obtain the NB distribution with the pmf

$$f(y;r,p)=\binom{y+r-1}{y}p^{r}(1-p)^{y}. \tag{6}$$

Corollary 3:

For $\alpha=\beta=1$, we obtain the Poisson-exponential or geometric distribution with the pmf

$$f(y;\theta)=\frac{\theta}{(\theta+1)^{y+1}}. \tag{7}$$

3. Some Properties of the PGL Distribution

This section presents some basic mathematical properties of the PGL distribution, specifically the moment generating function, probability generating function and the kth moment.

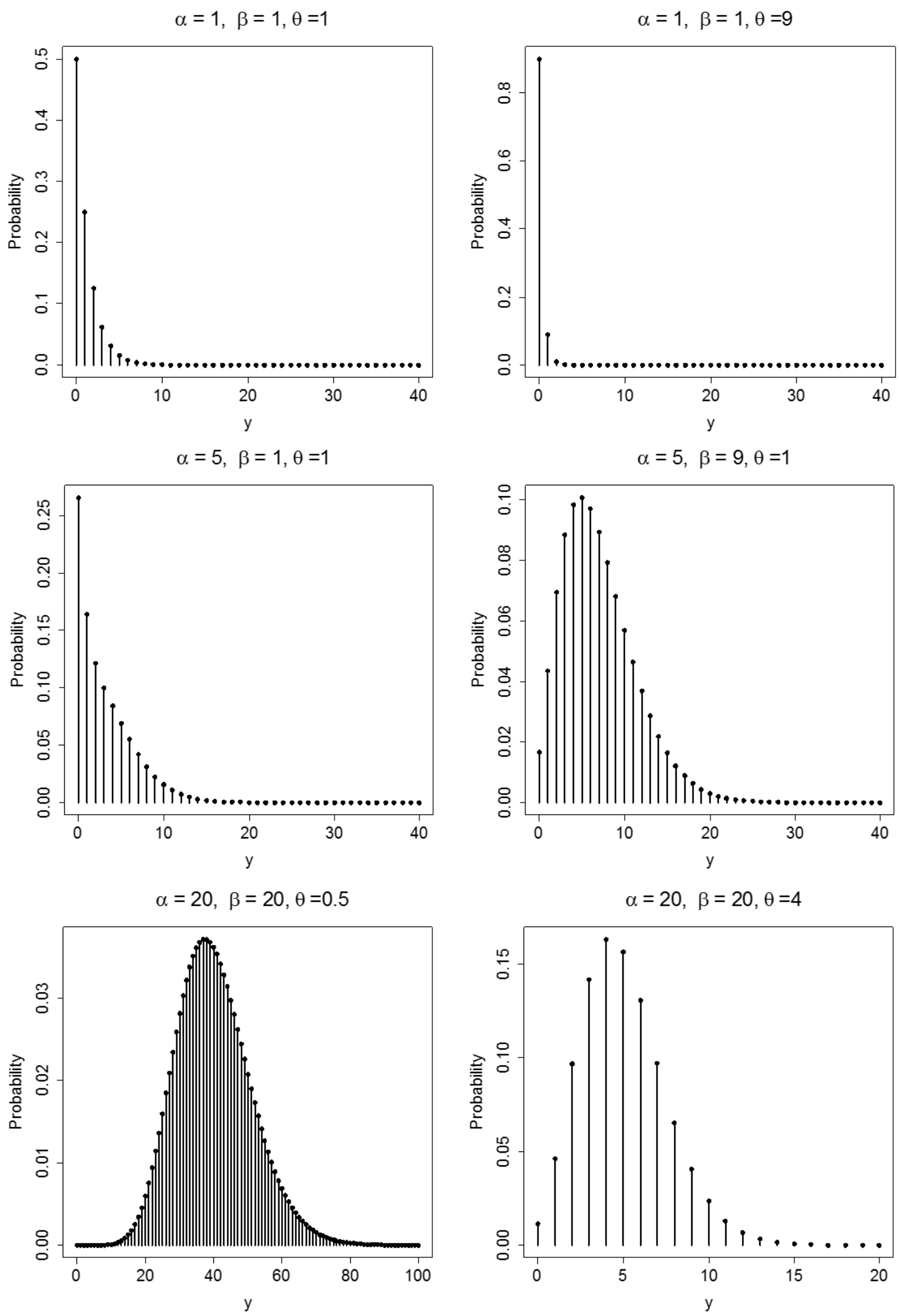

Figure 1. Some unimodal pmf plots of the PGL distribution with specified parameter values

3.1 Moment generating function

Proposition 2:

Let Y be a random variable with the PGL probability function, the moment generating function (mgf) of $Y \sim \text{PGL}(\alpha,\beta,\theta)$ is

$$M_Y(t) = \frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}}{(\theta-\exp(t)+1)^{\alpha}} + \frac{\theta^{\beta}}{(\theta-\exp(t)+1)^{\beta}}\right);$$

$\exp(t) < (\theta+1).$

Proof:

The mgf of mixed Poisson distribution can be obtained from

$$M_Y(t) = \text{E}(\exp(tY))$$

$$= \int_0^{\infty}\sum_{y=0}^{\infty}\exp(ty)\frac{\exp(-\lambda)\lambda^{y}}{y!}g(\lambda)\text{d}\lambda,$$

since $\sum_{y=0}^{\infty}\exp(ty)\exp(-\lambda)\lambda^{y}/y! = \exp(\lambda(\exp(t)-1))$ is the mgf of Poisson distribution, the mgf of PGL will be

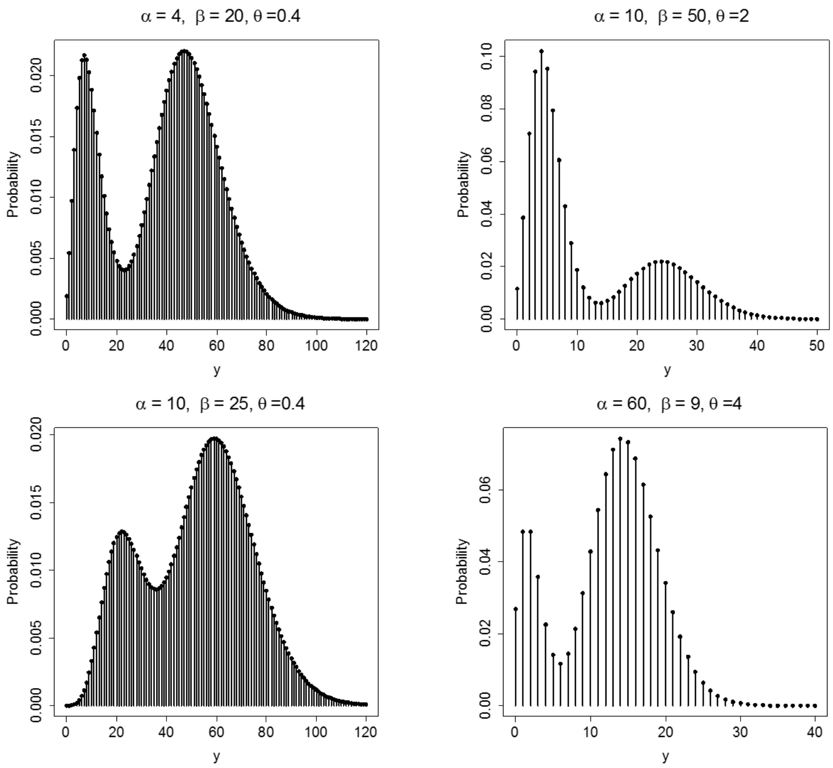

Figure 2. Some bimodal pmf plots of the PGL distribution with specified parameter values

$$M_Y(t)=\int_0^{\infty}\exp(\lambda(\exp(t)-1))g(\lambda)\mathrm{d}\lambda$$

$$=\int_0^{\infty}\exp(\lambda(\exp(t)-1))\frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}\lambda^{\alpha-1}}{\Gamma(\alpha)}+\frac{\theta^{\beta}\lambda^{\beta-1}}{\Gamma(\beta)}\right)\exp(-\theta\lambda)\mathrm{d}\lambda$$

$$=\frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}}{\Gamma(\alpha)}\int_0^{\infty}\lambda^{\alpha-1}\exp(-(\theta-\exp(t)+1)\lambda)\mathrm{d}\lambda+\frac{\theta^{\beta}}{\Gamma(\beta)}\int_0^{\infty}\lambda^{\beta-1}\exp(-(\theta-\exp(t)+1)\lambda)\mathrm{d}\lambda\right)$$

$$=\frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}}{(\theta-\exp(t)+1)^{\alpha}}+\frac{\theta^{\beta}}{(\theta-\exp(t)+1)^{\beta}}\right).$$

3.2 Probability generating function

Proposition 3:

Let Y be a random variable with the PGL probability function, the probability generating function (pgf) of $Y\sim \mathrm{PGL}(\alpha,\beta,\theta)$ is

$$H(s)=\frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}}{(\theta-s+1)^{\alpha+1}}+\frac{\theta^{\beta}}{(\theta-s+1)^{\beta+1}}\right);\ s<(\theta+1).$$

Proof:

The pgf of mixed Poisson distribution can be obtained by utilising the pgf of Poisson distribution as follows

$$H(s)=\mathrm{E}(s^Y)$$

$$=\int_0^{\infty}\sum_{y=0}^{\infty}s^y\frac{\exp(-\lambda)\lambda^y}{y!}g(\lambda)\mathrm{d}\lambda,$$

since $\sum_{y=0}^{\infty}s^y\exp(-\lambda)\lambda^y/y!=\exp(-(1-s)\lambda)$, it is the pgf of Poisson distribution, then the pgf of PGL will be

$$H(s)=\int_0^{\infty}\exp(-(1-s)\lambda)g(\lambda)\mathrm{d}\lambda$$

$$=\frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}}{\Gamma(\alpha)}\int_0^{\infty}\lambda^{\alpha-1}\exp(-(\theta-s+1)\lambda)\mathrm{d}\lambda+\frac{\theta^{\beta}}{\Gamma(\beta)}\int_0^{\infty}\lambda^{\beta-1}\exp(-(\theta-s+1)\lambda)\mathrm{d}\lambda\right)$$

$$=\frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}}{(\theta-s+1)^{\alpha}}+\frac{\theta^{\beta}}{(\theta-s+1)^{\beta}}\right).$$

Alternatively, the pgf of the PGL distribution can be got by setting $s=\exp(t)$ in the expression for the mgf.

3.3 Moments

Proposition 4:

Let Y be a random variable with the PGL probability function, the *k*th factorial moment of $Y \sim \mathrm{PGL}(\alpha,\beta,\theta)$ is

Proof:

The *k*th factorial moment of a mixed Poisson distribution can be found by

$$\mu'_{[k]} = \frac{1}{\theta+1}\left(\frac{\Gamma(k+\alpha)}{\Gamma(\alpha)\theta^{k-1}} + \frac{\Gamma(k+\beta)}{\Gamma(\beta)\theta^{k}}\right).$$

since $\sum_{y=0}^{\infty} y^k \exp(-\lambda)\lambda^y / y! = \lambda^k$, it is the *k*th moment about origin of the Poisson then we obtain

$$\mu'_{[k]} = \int_0^\infty \lambda^k g(\lambda)\mathrm{d}\lambda$$

$$= \int_0^\infty \lambda^k \frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}\lambda^{\alpha-1}}{\Gamma(\alpha)} + \frac{\theta^{\beta}\lambda^{\beta-1}}{\Gamma(\beta)}\right)\exp(-\theta\lambda)\mathrm{d}\lambda$$

$$= \frac{1}{\theta+1}\left(\frac{\theta^{\alpha+1}}{\Gamma(\alpha)}\int_0^\infty \lambda^{k+\alpha-1}\exp(-\theta\lambda)\mathrm{d}\lambda + \frac{\theta^{\beta}}{\Gamma(\beta)}\int_0^\infty \lambda^{r+\beta-1}\exp(-\theta\lambda)\mathrm{d}\lambda\right)$$

$$= \frac{1}{\theta+1}\left(\frac{\Gamma(k+\alpha)}{\Gamma(\alpha)\theta^{k-1}} + \frac{\Gamma(k+\beta)}{\Gamma(\beta)\theta^{k}}\right).$$

The *k*th moment about the mean is also called the *k*th central moment, $\mu_k = \mathrm{E}[(Y-\mu)^k] = \sum_{j=0}^{k}(-1)^j \binom{k}{j}\mu'_{k-j}\mu'^{j}$.

Consequently, the first four central moments of $Y \sim \mathrm{PGL}(\alpha,\beta,\theta)$ are

$$\mu_2 = \frac{\beta(\theta(-2\alpha+\theta+2)+1) + \alpha\theta(\alpha+(\theta+1)^2) + \beta^2\theta}{\theta^2(\theta+1)^2},$$

$$\mu_3 = \frac{3(\alpha\theta+\beta)^3}{\theta^3(\theta+1)^3} - \frac{3(\alpha\theta+\beta)\left(\alpha\theta+\alpha(\alpha+1)+\frac{\beta(\beta+1)}{\theta}+\beta\right)}{\theta^2(\theta+1)^2}$$
$$+\frac{1}{\theta+1}\left(\frac{\alpha(\alpha+1)(\alpha+2)}{\theta^2} + \frac{3\alpha(\alpha+1)}{\theta} + \alpha + \frac{\beta(\beta+1)(\beta+2)}{\theta^3} + \frac{3\beta(\beta+1)}{\theta^2} + \frac{\beta}{\theta}\right)$$

$$\mu_4 = -\frac{3(\alpha\theta+\beta)^4}{\theta^4(\theta+1)^4} + \frac{6\left(\alpha\theta+\alpha(\alpha+1)+\frac{\beta(\beta+1)}{\theta}+\beta\right)(\alpha\theta+\beta)^2}{\theta^3(\theta+1)^3}$$
$$-\frac{4\left(\frac{(\alpha+1)(\alpha+2)\alpha}{\theta^2} + \frac{3(\alpha+1)\alpha}{\theta} + \alpha + \frac{\beta(\beta+1)(\beta+2)}{\theta^3} + \frac{3\beta(\beta+1)}{\theta^2} + \frac{\beta}{\theta}\right)(\alpha\theta+\beta)}{\theta(\theta+1)^2}$$
$$+\frac{1}{\theta+1}\left(\frac{\alpha(\alpha+2)(\alpha+3)(\alpha+1)}{\theta^3} + \frac{6\alpha(\alpha+2)(\alpha+1)}{\theta^2} + \frac{7\alpha(\alpha+1)}{\theta}\right.$$

$$+\frac{\beta(\beta+1)(\beta+2)(\beta+3)}{\theta^4}+\frac{6\beta(\beta+1)(\beta+2)}{\theta^3}+\frac{7\beta(\beta+1)}{\theta^2}+\frac{\beta}{\theta}\Bigg).$$

In particular, the mean, variance, skewness and kurtosis of $Y \sim \text{PGL}(\alpha,\beta,\theta)$ according to its generating function, respectively, are

$$\text{E}(Y)=\frac{\alpha\theta+\beta}{\theta(\theta+1)},\ \ \text{Var}(Y)=\frac{\beta(\theta(-2\alpha+\theta+2)+1)+\alpha\theta(\alpha+(\theta+1)^2)+\beta^2\theta}{\theta^2(\theta+1)^2},$$

$$\text{Skewness}(Y)=\frac{\mu_3}{\mu_2^{3/2}},\text{ and Kurtosis}(Y)=\frac{\mu_4}{\mu_2^2}.$$

4. Parameter Estimation

A widely used method of estimating the parameters of a distribution is by maximising the log-likelihood function of parameters, ℓ, called maximum likelihood estimation (MLE). Let $Y_1,Y_2,\ldots,Y_n$ be an independent and identically distributed random variables which has the PGL distribution, and correspond to $y_1,y_2,\ldots,y_n$ which is a random sample from the PGL population. Let $\mathbf{\Theta}=(\alpha,\beta,\theta)^T$ be the vector of the parameters. The likelihood function of the PGL distribution is

$$L(\mathbf{\Theta})=\prod_{i=1}^{n}\frac{1}{y_i!(\theta+1)^{y_i+1}}\left(\left(\frac{\theta}{\theta+1}\right)^{\alpha}\frac{\theta\Gamma(y_i+\alpha)}{\Gamma(\alpha)}+\left(\frac{\theta}{\theta+1}\right)^{\beta}\frac{\Gamma(y_i+\beta)}{\Gamma(\beta)}\right).$$

We can write the log-likelihood function as

$$\ell(\mathbf{\Theta})=\sum_{i=1}^{n}\log\left(\theta^{\alpha+1}(\theta+1)^{\beta}\Gamma(\beta)\Gamma(y_i+\alpha)+\theta^{\beta}(\theta+1)^{\alpha}\Gamma(\alpha)\Gamma(y_i+\beta)\right)$$

$$-\sum_{i=1}^{n}\log y_i!-(\sum_{i=1}^{n}y_i+n)\log(\theta+1)-n\log\Gamma(\alpha)-n\log\Gamma(\beta)-n(\alpha+\beta)\log(\theta+1),$$

and the first partial derivatives of the log-likelihood with respect to each parameter, called the score functions, are

$$\frac{\partial\ell(\mathbf{\Theta})}{\partial\alpha}=\sum_{i=1}^{n}\Bigg(\frac{\theta^{\alpha+1}(\theta+1)^{\beta}\Gamma(y_i+\alpha)\Gamma(\beta)(\psi(y_i+\alpha)+\log\theta)}{\theta^{\alpha+1}(\theta+1)^{\beta}\Gamma(\beta)\Gamma(y_i+\alpha)+\theta^{\beta}(\theta+1)^{\alpha}\Gamma(\alpha)\Gamma(y_i+\beta)}$$

$$+\frac{\theta^{\beta}(\theta+1)^{\alpha}\Gamma(\alpha)\Gamma(y_i+\beta)(\psi(\alpha)+\log(\theta+1))}{\theta^{\alpha+1}(\theta+1)^{\beta}\Gamma(\beta)\Gamma(y_i+\alpha)+\theta^{\beta}(\theta+1)^{\alpha}\Gamma(\alpha)\Gamma(y_i+\beta)}\Bigg)$$

$$-n\psi(\alpha)-n\log(\theta+1),$$

$$\frac{\partial\ell(\mathbf{\Theta})}{\partial\beta}=\sum_{i=1}^{n}\Bigg(\frac{\theta^{\alpha+1}(\theta+1)^{\beta}\Gamma(y_i+\alpha)\Gamma(\beta)(\psi(\beta)+\log(\theta+1))}{\theta^{\alpha+1}(\theta+1)^{\beta}\Gamma(\beta)\Gamma(y_i+\alpha)+\theta^{\beta}(\theta+1)^{\alpha}\Gamma(\alpha)\Gamma(y_i+\beta)}$$

$$+\frac{\theta^{\beta}(\theta+1)^{\alpha}\Gamma(\alpha)\Gamma(y_i+\beta)(\psi(y_i+\beta)+\log\theta)}{\theta^{\alpha+1}(\theta+1)^{\beta}\Gamma(\beta)\Gamma(y_i+\alpha)+\theta^{\beta}(\theta+1)^{\alpha}\Gamma(\alpha)\Gamma(y_i+\beta)}\Bigg)$$

$$-n\psi(\beta)-n\log(\theta+1),$$

and

$$\frac{\partial\ell(\mathbf{\Theta})}{\partial\theta}=\sum_{i=1}^{n}\Bigg(\frac{\Gamma(\beta)\Gamma(y_i+\alpha)\left(\theta^{\alpha}(\theta+1)^{\beta-1}(\beta\theta+(\alpha+1)(\theta+1))\right)}{\theta^{\alpha+1}(\theta+1)^{\beta}\Gamma(\beta)\Gamma(y_i+\alpha)+\theta^{\beta}(\theta+1)^{\alpha}\Gamma(\alpha)\Gamma(y_i+\beta)}$$

$$+\frac{\Gamma(\alpha)\Gamma(y_i+\beta)\left(\theta^{\beta-1}(\theta+1)^{\alpha-1}(\alpha\theta+\beta(\theta+1))\right)}{\theta^{\alpha+1}(\theta+1)^{\beta}\Gamma(\beta)\Gamma(y_i+\alpha)+\theta^{\beta}(\theta+1)^{\alpha}\Gamma(\alpha)\Gamma(y_i+\beta)}$$

$$-\left(\frac{(\sum_{i=1}^{n}y_i+n)+n(\alpha+\beta)}{\theta+1}\right),$$

where $\psi(t)=\frac{\mathrm{d}}{\mathrm{d}t}\log\Gamma(t)$ is the digamma function.

The maximum likelihood estimators of the PGL distribution can be achieved by setting the score functions equal to zero, giving the so-called maximum likelihood score equations, and solving this system of equations. In this case, the score functions are nonlinear and do not have analytical solution. Instead, maximum likelihood estimates can be obtained by a numerical method (e.g., Newton-Raphson method, Nelder-Mead method, BFGS method, SANN method, as implemented in an R function `mle2`).

5. Applications to Real Data Sets

Some real data sets are considered to fit with the proposed distribution (PGL), Poisson, NB and PL distributions. The first data set is number of the mistakes in copying groups of random digits that was used for illustrating the PL distribution by Sankaran (1970). The second data set is the number of micronuclei after exposure at dose 4 (Gy) of γ - Irradiation. They were counted using the cytochalasin B method and fitted with the NB distribution (Puig and Valero, 2006). The third data set is an application in genetics, the number of chromatid aberrations (0.2 g chinon 1, 24 hours). It had been fitted previously with the Poisson and the PL distributions, but given the amount of over-dispersion in the data, the PL distribution is a more appropriate model (Shanker and Fesshaye, 2015).

Another application involving bimodal data is also considered in this part. The data set is the number of Chemopodium album in arable land per quadrat, which was fit with the NB distribution (Bliss and Fisher, 1953). We fit this data set with the proposed distribution, the NB distribution and a five-parameter mixture of two NB distributions (MixtureNB) with the weighted parameter ω, where $0\le\omega\le 1$, with pmf

$$f(y;r_1,r_2,p_1,p_2,\omega)=\omega\frac{\Gamma(y+r_1)}{y!\Gamma(r_1)}p_1^{r_1}\left(1-p_1\right)^{y}+(1-\omega)\frac{\Gamma(y+r_2)}{y!\Gamma(r_2)}p_2^{r_2}\left(1-p_2\right)^{y}.$$

Descriptive summaries of these data are shown in Table 1. The index of dispersion is greater than unity for all data sets, indicating that all data sets are over-dispersed.

In this work, the SANN method based on bbmle package (Bolker and Team, 2014) of the R programming language (R Core Team, 2014), being a global optimization, is used to compute the maximum likelihood estimates (Nash, 2014).

Tables 2, 3, 4 and 5 present the results of fitting the different distributions to these real data sets. We use the estimated log-likelihood (LL) and Anderson-Darling (AD) test for discrete distributions to compare the expected and observed values of each data set. The AD-test is an empirical distribution function goodness-of-fit test for discrete data (Choulakian *et al.*, 1994).

The null hypothesis is that data follow whatever distribution that is being tested including Poisson, NB, PL, MixtureNB, and PGL with given parameter estimates against the alternative that data follow some other distributions. The discrete AD-test can be obtained by using `dgof` package (Arnold and Emerson, 2011) of the R programming language.

Table 1. Summary data

	Min	Mode	Max	Mean	Dispersion
Number of mistakes in copying groups	0	0	4	0.7833	1.6051
Number of micronuclei	0	0	7	1.0132	1.1725
Number of chromatid aberrations	0	0	7	0.5475	2.0558
Number of Chenopodium album per quadrat	0	5	10	4.0316	1.9551

Table 2. The number of mistakes in copying groups of random digits

The mistakes in copying groups	Observed values	Expected values			
		Poisson	NB	PL	PGL
0	35	27.41	33.97	33.06	34.49
1	11	21.47	14.51	15.27	12.65
2	8	8.41	6.39	6.74	7.03
3	4	2.2	2.84	2.88	3.36
4	2	0.43	1.27	1.21	1.47
Parameter estimates		$\hat{\lambda}=0.7833$	$\hat{r}=0.9421$ $\hat{p}=0.5456$	$\hat{\theta}=1.7434$	$\hat{\alpha}=2.7084$ $\hat{\beta}=0.0039$ $\hat{\theta}=2.3928$
LL		-77.5456	-73.3684	-73.3510	-72.5825
AD-statistic		2.2733	0.1546	0.2287	0.0518
***p*-value**		0.0494	0.8286	0.7395	0.9680

Table 3. The number of micronuclei

The number of micronuclei	Observed values	Expected values			
		Poisson	NB	PL	PGL
0	1974	1816.04	1965.37	2396.75	1975.37
1	1674	1839.97	1695.41	1300.33	1676.23
2	869	932.11	857.66	668.83	863.87
3	342	314.8	331.87	332.16	336.85
4	102	79.74	108.68	160.92	108.87
5	26	16.16	31.71	76.53	30.69
6	13	2.73	8.5	35.88	7.79
7	2	0.39	2.10	16.63	1.73
Parameter estimates		$\hat{\lambda}=1.0132$	$\hat{r}=5.8154$ $\hat{p}=0.8517$	$\hat{\theta}=1.3873$	$\hat{\alpha}=9.22$ $\hat{\beta}=2.9427$ $\hat{\theta}=8.4507$
LL		-6767.9100	-6735.9057	-6918.3639	-6735.7035
AD-statistic		10.664	0.1000	64.3591	0.0221
***p*-value**		0.0003	0.9545	0.0000	0.9985

The fitted distributions for the number of mistakes in copying groups are shown in Table 2. It illustrates that the PGL distribution gives the largest LL value. Although, the differences between LL values are small, but the distances from the observed to expected values and the *p*-value based on the discrete AD-test indicate that the null hypothesis cannot be rejected at the 0.05 significant level. It verifies that the mistakes in copying groups follows the PGL distribution with the highest *p*-value and can model this data well.

The number of micronuclei are fitted. From the result in Table 3, the LL values from the NB and the PGL distributions are very similar. However, the expected values from the PGL distribution are very close to the observed values, resulting in the null hypothesis being accepted at the 0.05 level of significance with *p*-value 0.9985.

Fitting the distributions to the number of chromatid aberrations data set shows that the PGL distribution gives the largest value of LL (Table 4). Comparing the observed and expected values demonstrates that the PGL distribution again provides a good fit to the number of chromatid aberrations, with the highest *p*-value (0.9362).

In the case of bimodal data, the MixtureNB distribution seems to provide a bit more appropriate for the number of Chenopodium alblum data set. Based on -*p*-value, it indi-

Table 4. The number of chromatid aberrations (0.2 g chinon 1, 24 hours)

The number of chromatid aberrations	Observed values	Expected values			
		Poisson	NB	PL	PGL
0	268	231.36	270.34	257.02	264.83
1	87	126.67	78.53	93.39	91.7
2	26	34.67	29.79	32.76	23.54
3	9	6.33	12.18	11.21	8.78
4	4	0.87	5.16	3.77	5.09
5	2	0.09	2.23	1.25	3.07
6	1	0.01	0.98	0.41	1.66
7	3	0	0.43	0.13	0.79
Parameter estimates		$\hat{\lambda}=0.5475$	$\hat{r}=0.6205$ $\hat{p}=0.5318$	$\hat{\theta}=2.3804$	$\hat{\alpha}=4.7909$ $\hat{\beta}=42.1508$ $\hat{\theta}=13.3789$
LL		-439.5136	-399.8572	-403.455	-398.0406
AD-statistic		8.7108	0.1891	0.8576	0.0712
***p*-value**		0.0014	0.7586	0.2585	0.9362

Table 5. The number of Chenopodium album per quadrat

The number of Chenopodium album per quadrat	Observed values	Expected values		
		NB	MixtureNB	PGL
0	19	8.99	18.99	17.66
1	5	13.41	5.01	4.39
2	6	14.24	5.81	7.19
3	9	13.07	9.63	10.59
4	5	11.07	12.42	12.42
5	20	8.89	12.84	12.13
6	14	6.89	11.05	10.27
7	8	5.19	8.16	7.73
8	4	3.84	5.27	5.27
9	3	2.79	3.02	3.31
10	2	2	1.56	1.93
Parameter estimates		$\hat{r}=2.3648$ $\hat{p}=0.3689$	$\hat{r}_1=821.4177$ $\hat{p}_1=0.9998$ $\hat{r}_2=32751.02$ $\hat{p}_2=0.9998$	$\hat{\omega}=0.2278$ $\hat{\alpha}=21.5688$ $\hat{\beta}=0.3477$ $\hat{\theta}=4.2322$
LL		-233.0949	-212.7645	-214.5600
AD-statistic		4.1169	0.4291	0.6248
***p*-value**		0.0067	0.6987	0.5167

cates that the data follow the mixture of two NB distributions at the 0.05 level of significance. Due to the expense of 2 extra parameters of the MixtureNB distribution, the PGL distribution with close LL value can be chosen as a simpler model for fitting this data set.

Figure 3 shows plot of the observed values and the expected values related to those shown in Tables 2-5 of the PGL distribution. It illustrates that real data are very close to the PGL distribution. Therefore, the PGL distribution can be an alternative model for count data in some situations.

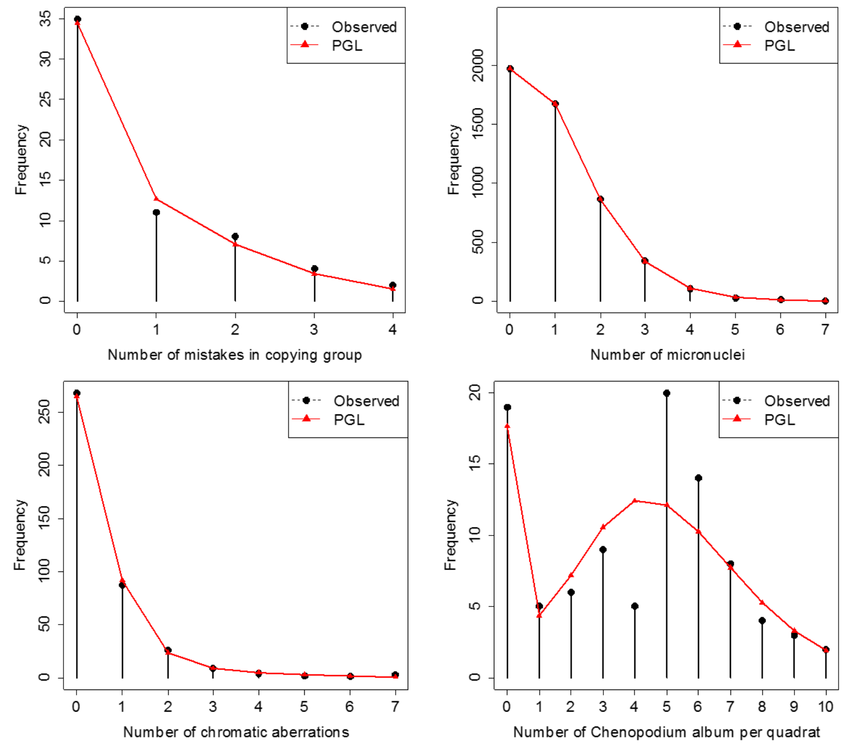

Figure 3. Results of fitting distributions to real data sets

6. Conclusions

In this work, a new mixed Poisson distribution is introduced. We consider that the mean of Poisson variable is an independent and identically distributed random variable according to a mixing distribution, a new generalised Lindley distribution. The proposed distribution is called the Poisson-generalised Lindley distribution. We have determined various mathematical properties of the Poisson-generalised Lindley variable, for instance, the probability mass function, moment generating function, probability generating function, the mean, and the variance. We show that the negative binomial, Poisson-Lindley, and Poisson-exponential distributions are special cases of it.

The proposed distribution is applied to several real data sets. The results, including the p-value based on the discrete Anderson-Darling test, indicate that the Poisson-generalised Lindley distribution is a flexible model that may be a useful alternative to other distributions for count data analysis.

Acknowledgements

The authors would like to thank to Department of Statistics, Faculty of Science, and the graduate school of Kasetsart University. Also we thank to Science Achievement Scholarship of Thailand (SAST) for supporting the first author.

References

Arnold, T. A. and Emerson, J. W. 2011. Nonparametric goodness-of-t tests for discrete null distributions. The R Journal. 3(2), 34-39.

Bliss, C. I. and Fisher, R. A. 1953. Fitting the negative binomial distribution to biological data. Biometrics. 9(2), 176-200.

Bolker, B. and Team, R. D. C. 2014. bbmle: Tools for general maximum likelihood estimation. R package version 1.0.17.

Choulakian, V., Lockhart, R. A., and Stephens, M. A. 1994. Cramér-von mises statistics for discrete distributions. Canadian Journal of Statistics. 22(1), 125-137.

Elbatal, I., Merovci, F., and Elgarhy, M. 2013. A new generalized Lindley distribution. Mathematical Theory and Modeling. 3, 30-47.

Everitt, B. and Hand, D. 1981. Finite mixture distributions. Monographs on Statistics and Applied Probability, Chapman and Hall, U.K.

Greenwood, M. and Yule, G. U. 1920. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. Journal of the Royal Statistical Society. 83(2), 255-279.

Holla, M. 1967. On a Poisson-inverse Gaussian distribution. Metrika. 11(1), 115-121.

Johnson, N. L., Kemp, A. W., and Kotz, S. 2005. Univariate Discrete Distributions, 3rd, Wiley Series in Probability and Statistics, John Wiley and Sons, Inc. Hoboken, New Jersey, U.S.A.

Karlis, D. and Xekalaki, E. 2005. Mixed Poisson distributions. International Statistical Review / Revue Internationale de Statistique. 73(1), 35-58.

Mahmoudi, E. and Zakerzadeh, H. 2010. Generalized Poisson - Lindley distribution. Communications in Statistics - Theory and Methods. 39(10), 1785-1798.

Nash, J. C. 2014. On best practice optimization methods in R. Journal of Statistical Software. 60(2), 1-14.

Panjer, H. H. 2006. Mixed Poisson Distributions. In Encyclopedia of Actuarial Science, John Wiley and Sons, Ltd. Hoboken, New Jersey, U.S.A.

Puig, P. and Valero, J. 2006. Count data distribution: Some characterizations with applications. Journal of the American Statistical Association. 101(473), 332-340.

R Core Team. 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Raghavachari, M., Srinivasan, A., and Sullo, P. 1997. Poisson mixture yield models for integrated circuits: A critical review. Microelectronics Reliability. 37(4), 565-580.

Rodríguez-Avi, J., Conde-Sanchéz, A., Saéz-Castillo, A. J., Olmo-Jimenéz, M. J., and Martínez- Rodríguez, A. M. 2009. A generalized Waring regression model for count data. Computational Statistics and Data Analysis. 53(10).

Sankaran, M. 1970. The discrete Poisson-Lindley distribution. International Biometric Society. 26(1). 145-149.

Shanker, R. and Fesshaye, H. 2015. On Poisson-Lindley distribution and its applications to biological sciences. Biometrics and Biostatistics International Journal. 2(4), 1-5.

Zamani, H., Ismail, N., and Faroughi, P. 2014. Poisson-weighted exponential univariate version and regression model with applications. Journal of Mathematics and Statistics. 10(2), 148-154.